Gene Regulatory Networks: Lessons from plants

Gabriel Krouk
November 9, 2017

UMR5004 CNRS Montpellier

Plan

1- Intro

2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

4- Is Regine cool? (REGINE)

Plan

1- Intro

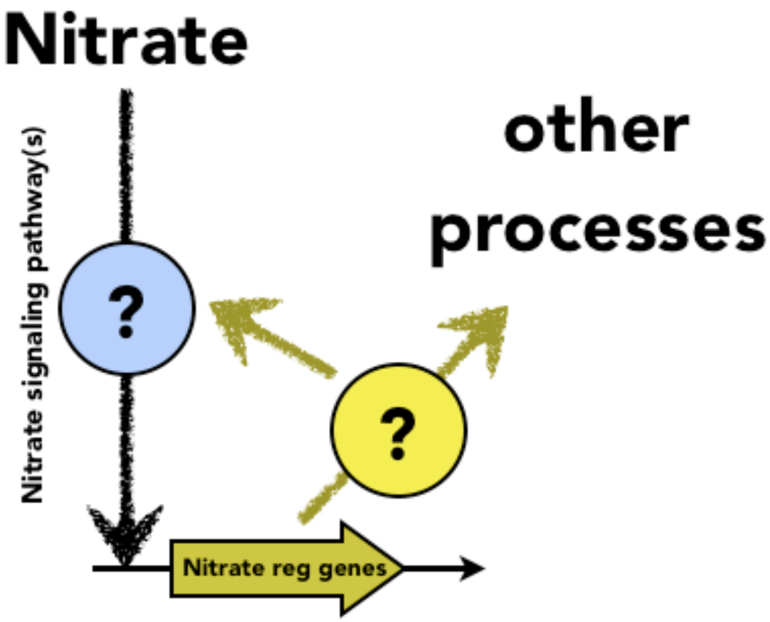
2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

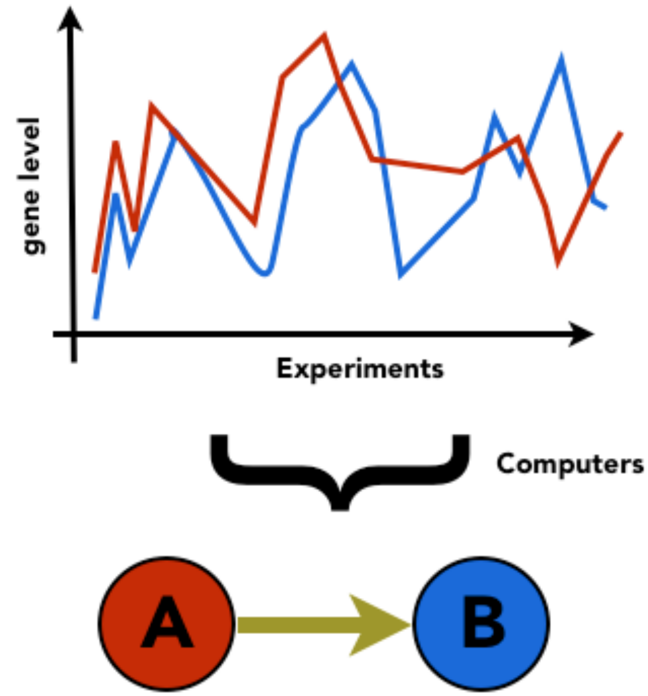
4- Is Regine cool? (REGINE)

Our questions

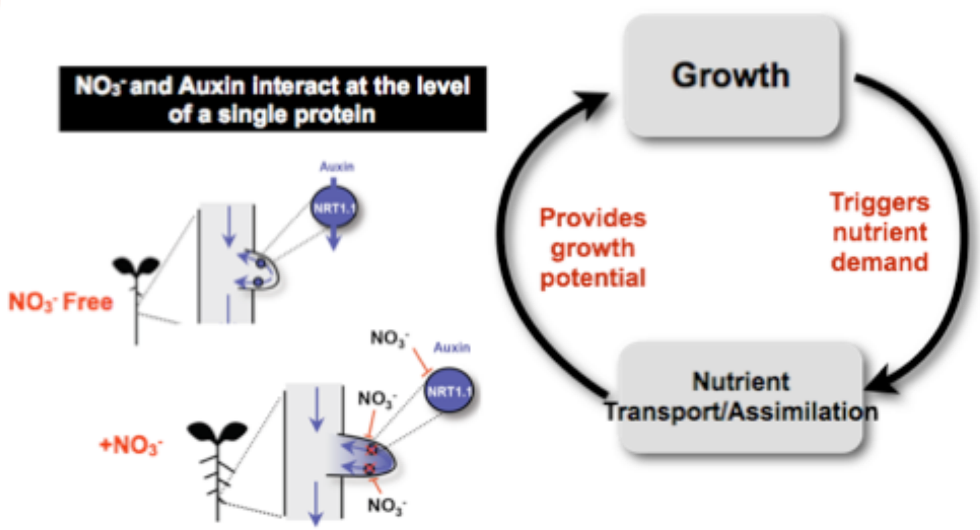
1)



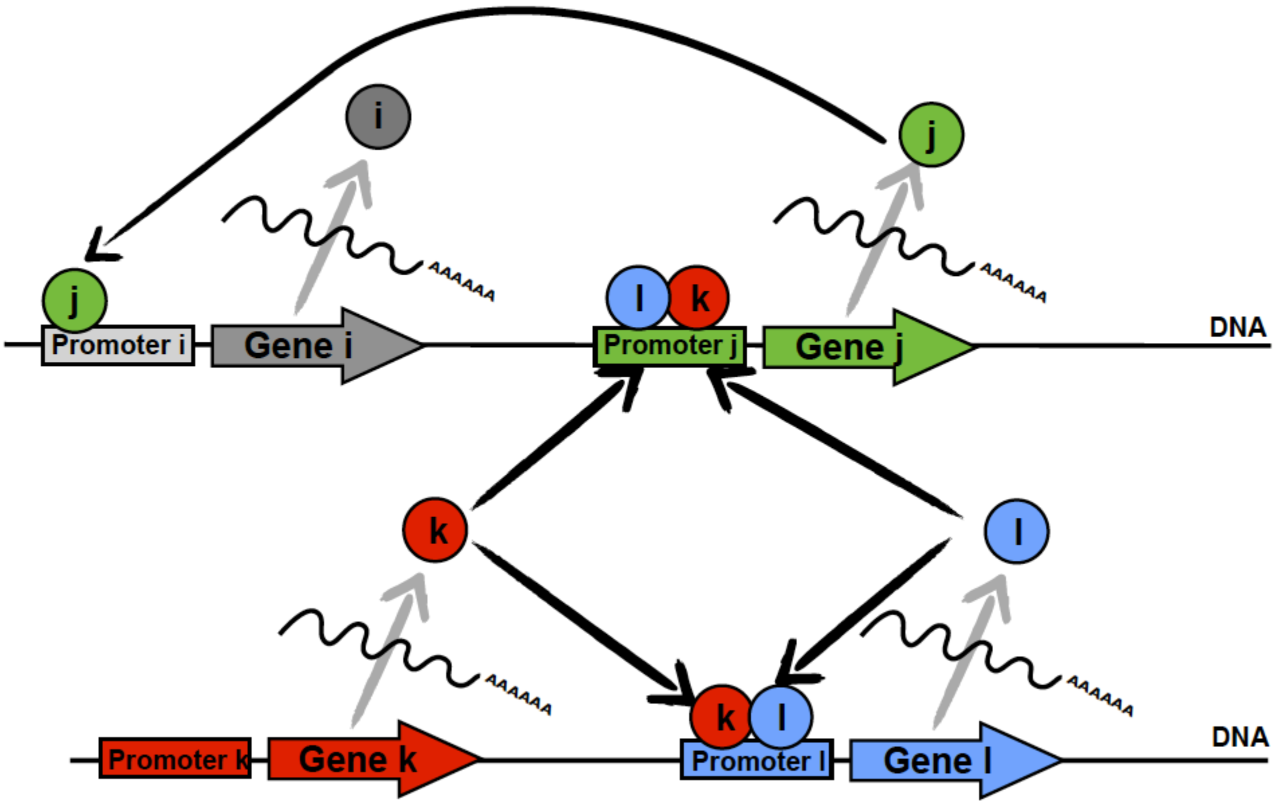
3)

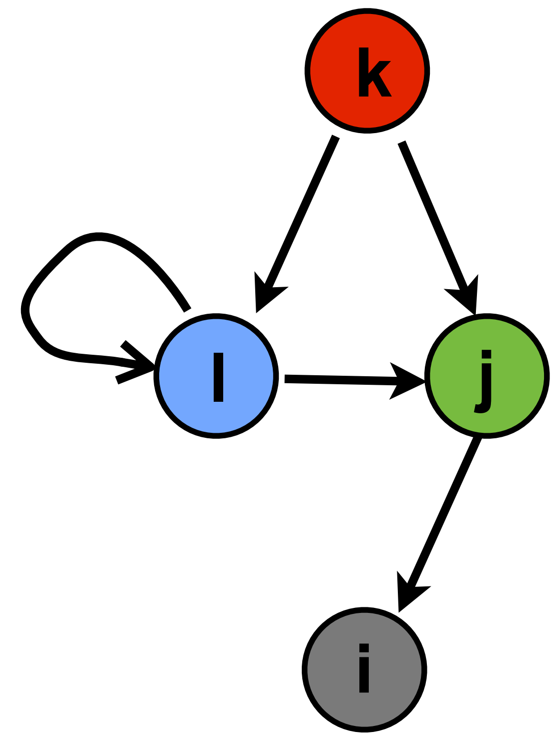


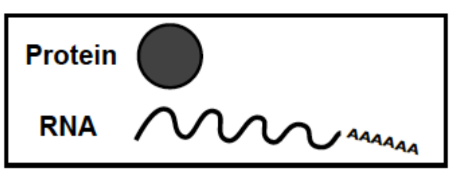
2)



For the geeks, and admin, around here.
Gene Regulatory Networks, qu'es aquò*?







Exhaustive vision of mRNA levels in cells.
microarrays, NGS etc…

de l'occitan: *qu'est ce que c'est?*

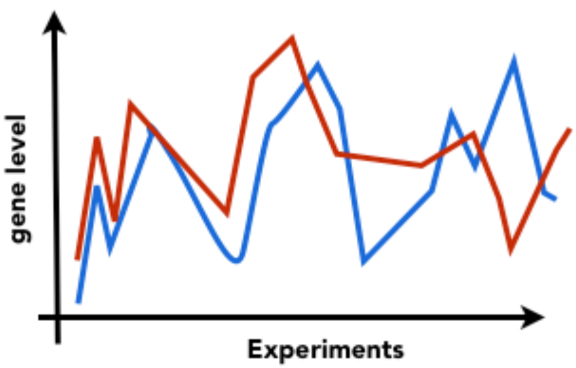
One particular important goal of Systems Biology:

Impacts may range from agronomy to medicine…

Gene Expression

Algorithms

GRNs



Experimental validation

Plan

1- Intro

2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

4- Is Regine cool?
(REGINE)

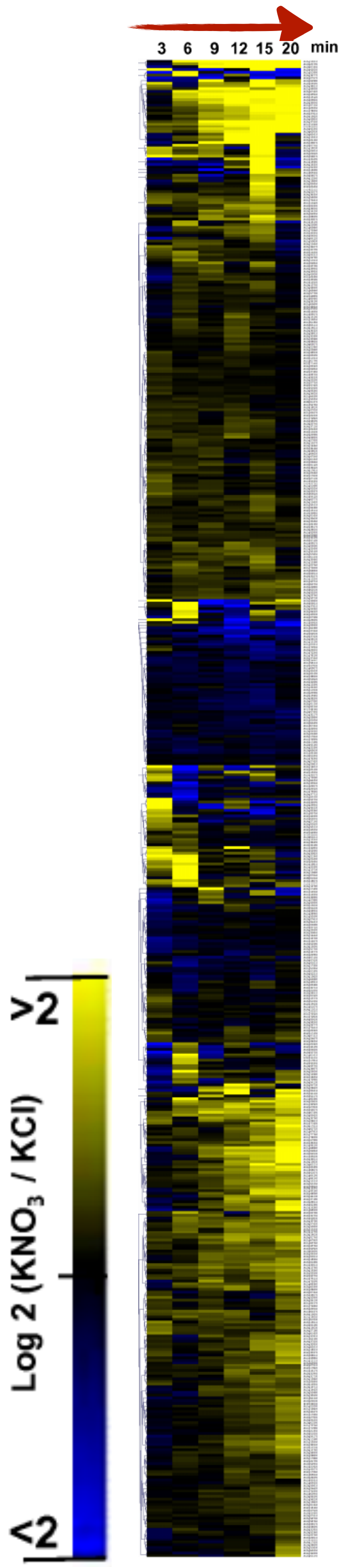
Plan

1- Intro

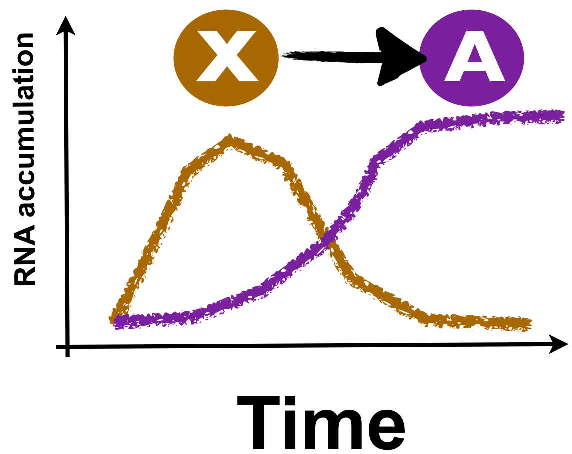
2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

4- Is Regine cool? (REGINE)



Minimal view of the Network



State Space Modeling

Dennis Shasha   Piotr Mirowski

Krouk *et al. Genome Biology* 2010, **11**:R123
http://genomebiology.com/content/11/12/R123

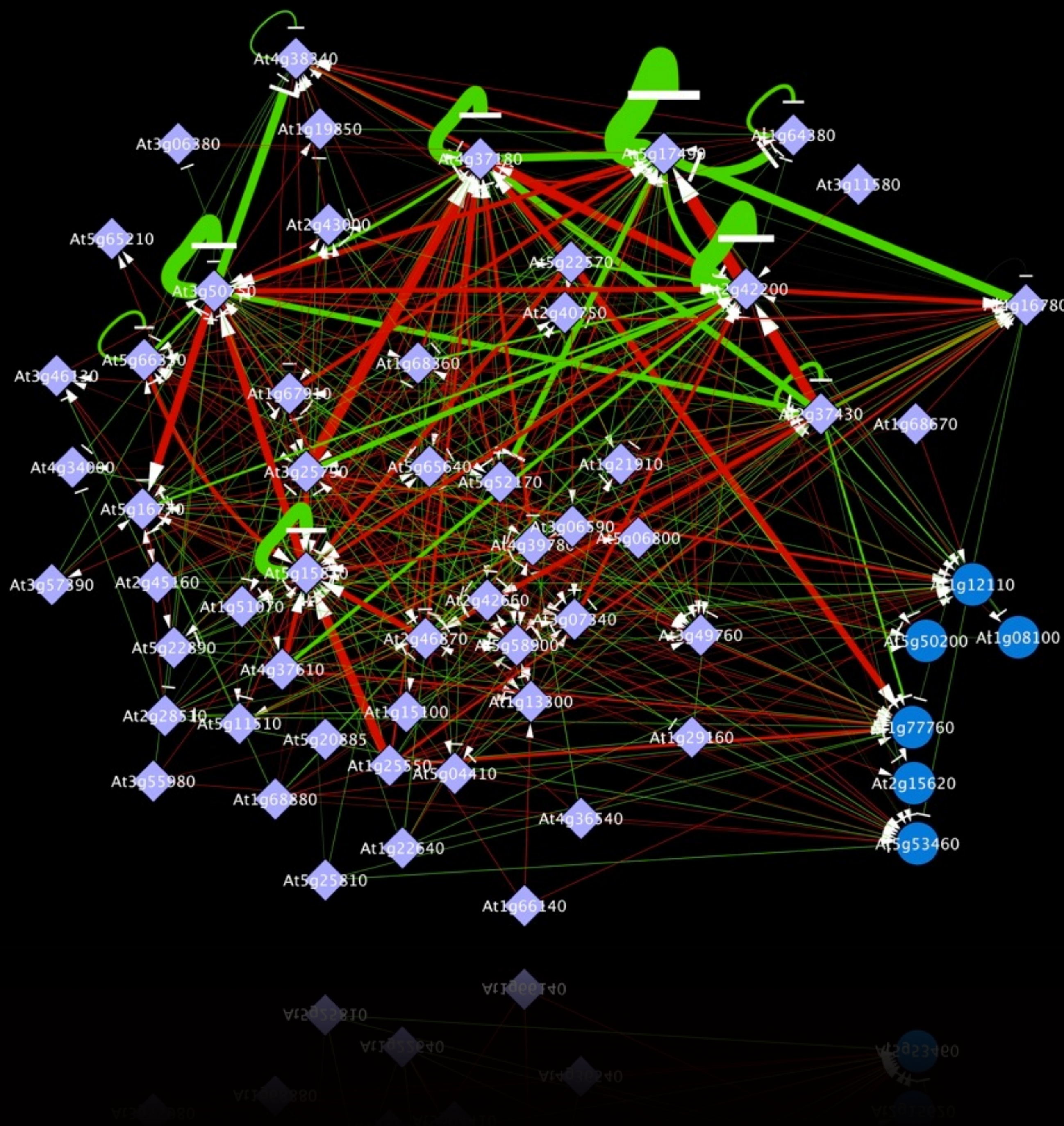
Genome **Biology**

RESEARCH **Open Access**

Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate

Gabriel Krouk[1,2], Piotr Mirowski[3], Yann LeCun[3], Dennis E Shasha[3], Gloria M Coruzzi[1*]

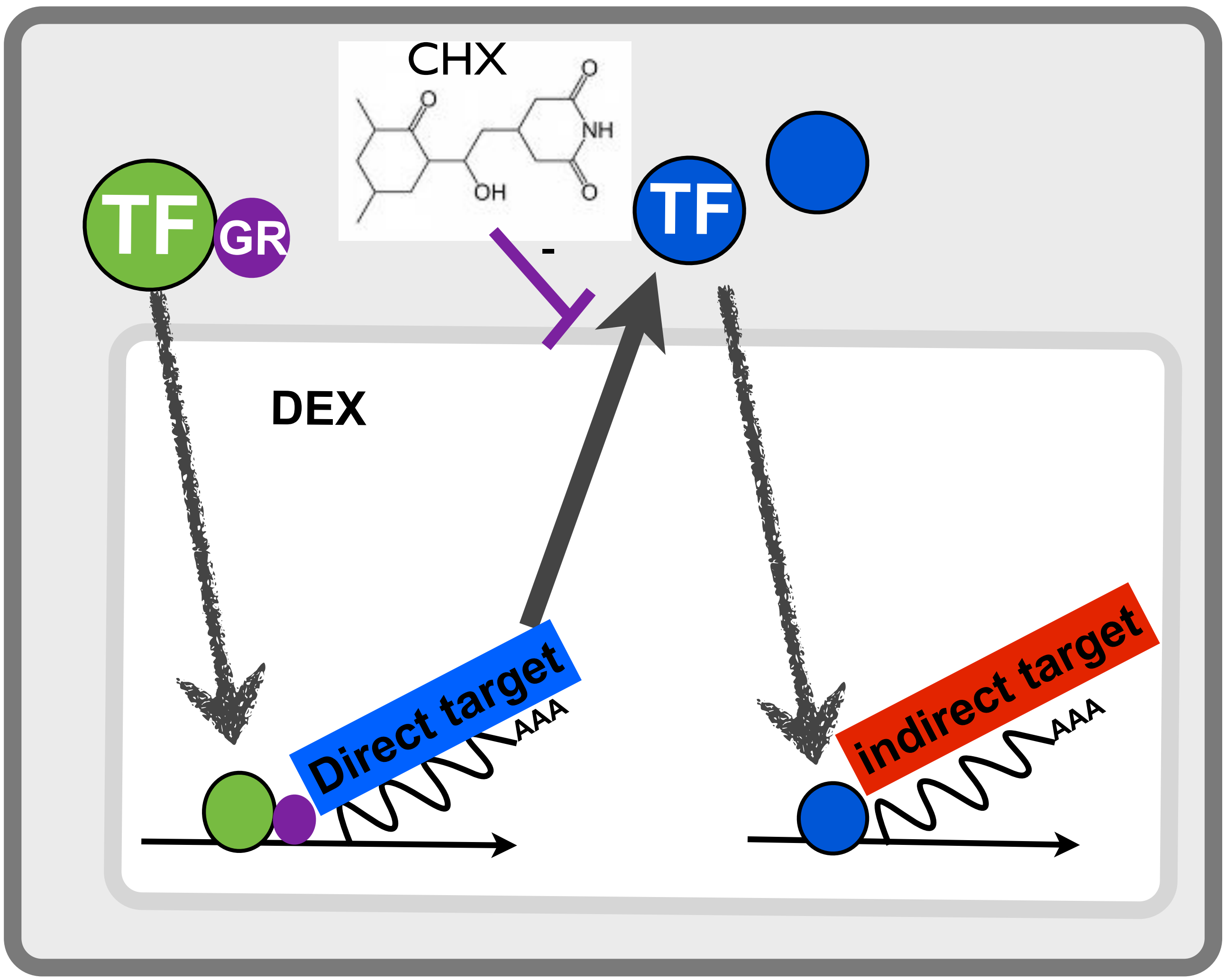
This figure looks crazy ... but …

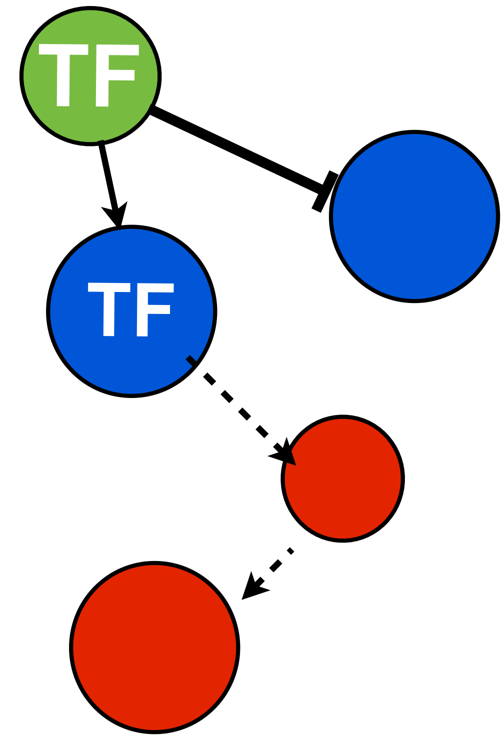


TARGET: Transient **A**ssay **R**eporting **G**enome-wide **E**ffects of **T**ranscription factors

the concept

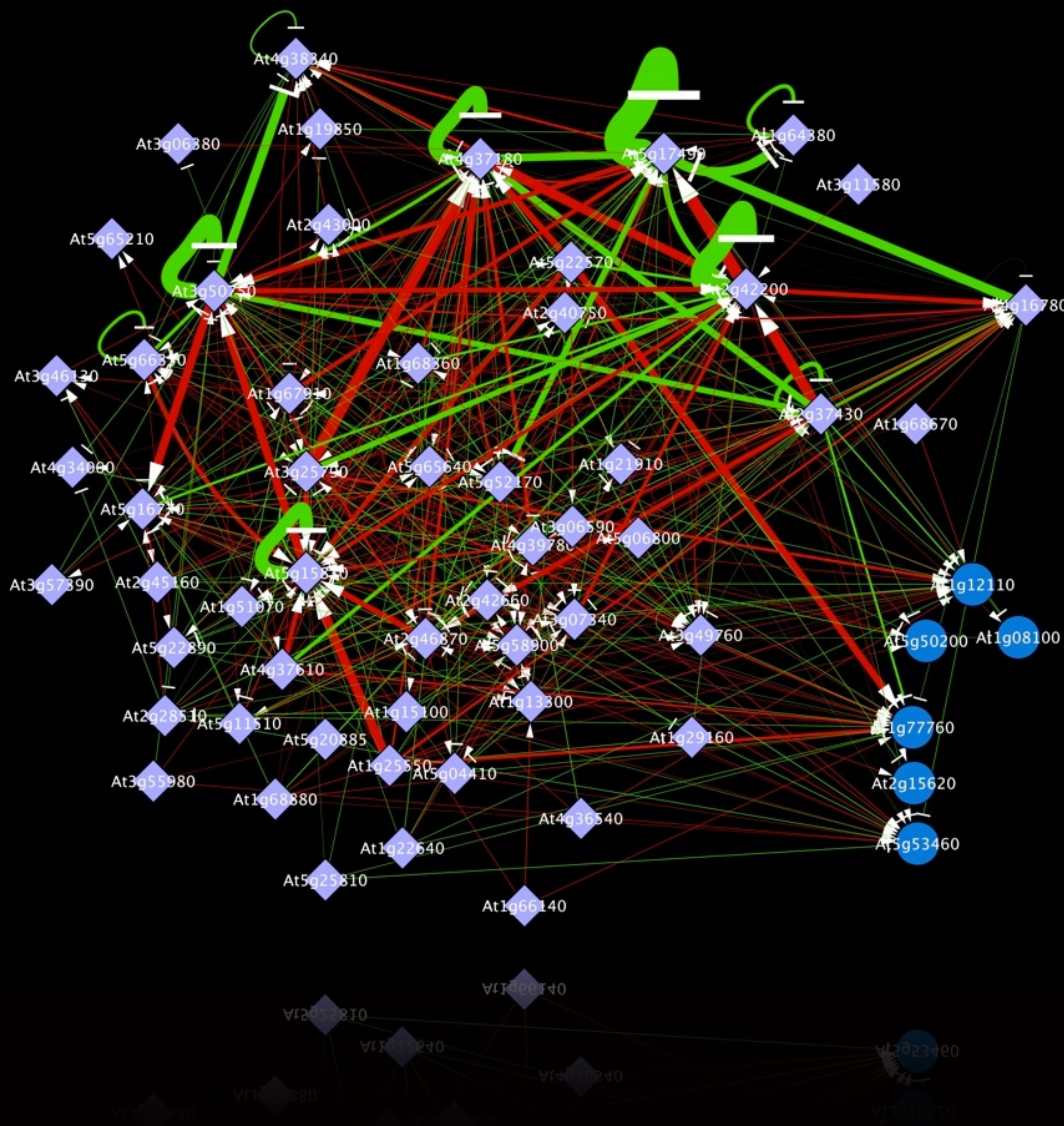
Adapted from Sablowski and Meyerowitz Cell, Vol. 92, 93–103, January 9, 1998





Bargmann et al. Mol Plant 2013
Para et al. PNAS 2014
Medici et al. Nature Commun 2015

this figure looks crazy ... but …



this figure looks crazy ... but …
it is possibly close to reality (connection-wise)

-We have many ways to experimentally
probe connections in GRNs (ie:Chip-Seq,TARGET, eY1H):

TARGET: One particular TF may control ~100 to 400 genes.

Knowing that ~1800-2000 TFs in Arabidopsis...

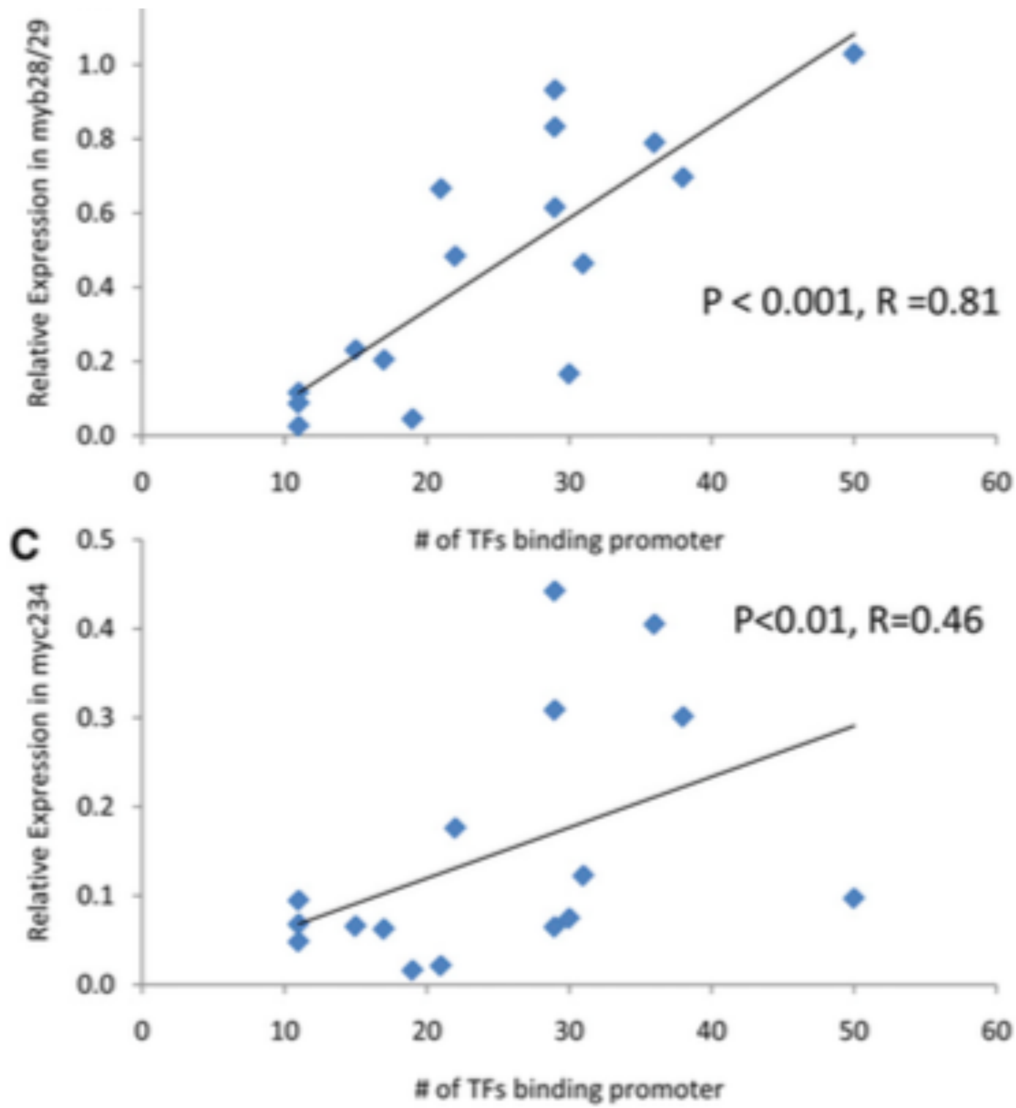
… in average a gene is controlled by

~6 to 40 TFs.

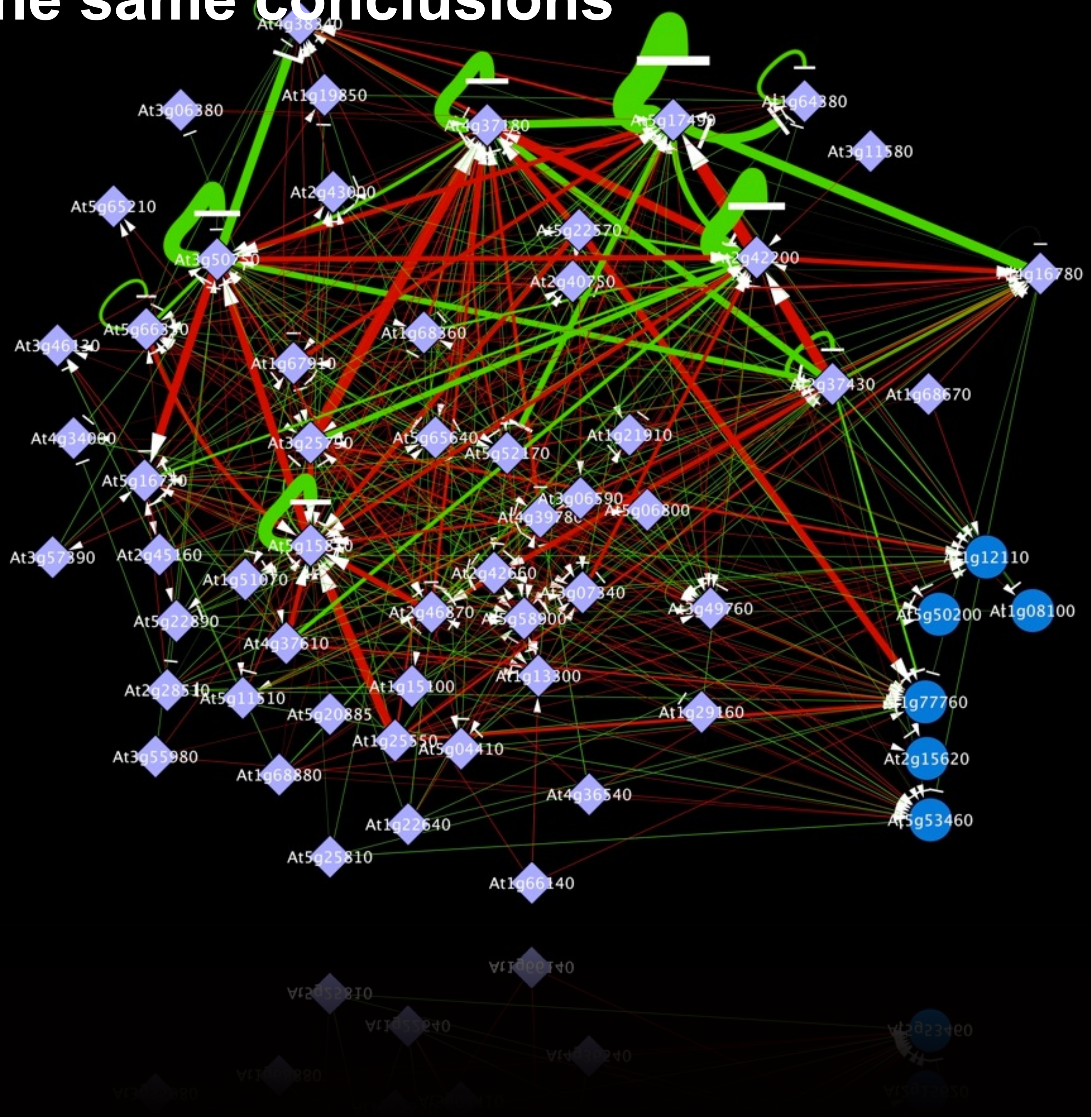
Promoter-Based Integration in Plant Defense Regulation[1][W][OPEN]

Baohua Li, Allison Gaudinier, Michelle Tang, Mallorie Taylor-Teeples, Ngoc T. Nham, Cyrus Ghaffari, Darik Scott Benson, Margaret Steinmann, Jennifer A. Gray, Siobhan M. Brady, and Daniel J. Kliebenstein*

Departments of Plant Sciences (B.L., M.T., N.T.N. C.G., D.S.B., M.S., J.A.G., D.J.K.) and Plant Biology (A.G., M.T., M.T.-T., J.A.G., S.M.B.) and Genome Center (A.G., M.T., M.T.-T., J.A.G., S.M.B.), University of California, Davis, California 95616; and DynaMo Center of Excellence, University of Copenhagen, DK–1871 Frederiksberg C, Denmark (D.J.K.)



Machine learning and Experimental procedures converge towards the same conclusions

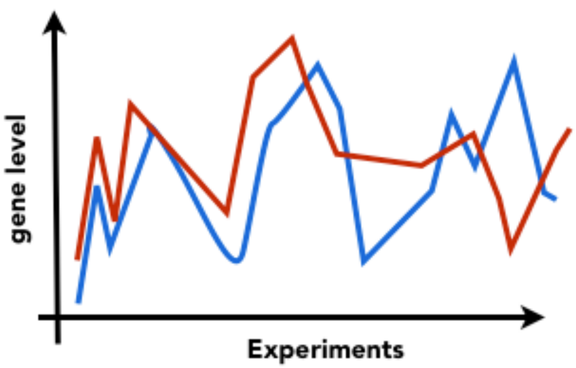


One particular important goal of Systems Biology:

GRNs

Gene Expression

Machine Learning



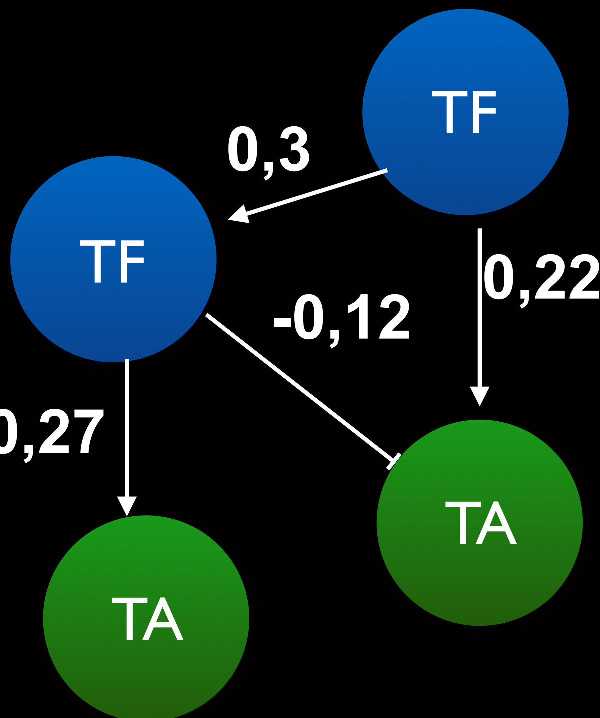
Experimental validation

Project- We use these parameters to simulate large ‘eukaryotic-like’ GRNs

Clement Carre Post-doc

Pr Andre Mas

Fast **R**andomizing **A**lgorithm for **N**etwork **K**nowledge

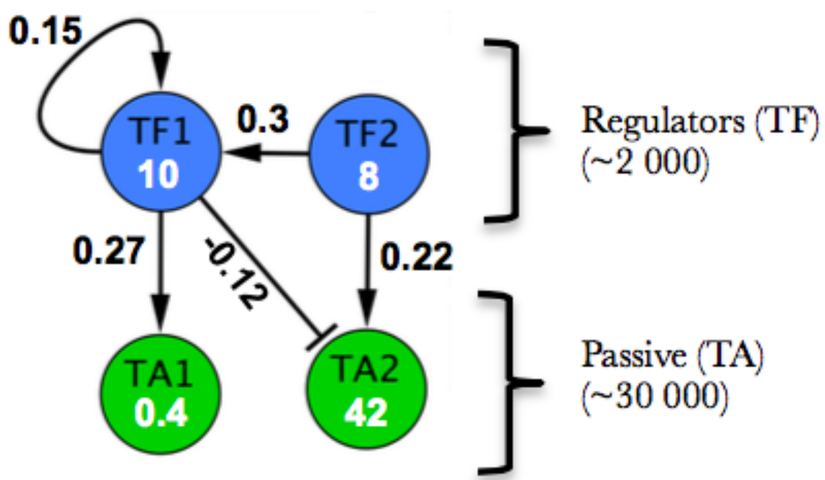


Regulators	2,000
Passive	30,000

Poses fundamental/mathematical questions concerning network stability and more…

Carré et al, *npj Systems Biology & App* (2017)

The Formalism = FRANK's gut is simple though highly scalable



Network (N)

$$\begin{array}{c} \\ \textbf{TF1} \\ \textbf{TF2} \end{array} \begin{array}{c} \begin{array}{cc} \textbf{TF1} & \textbf{TF2} \end{array} \\ \begin{bmatrix} 0.15 & 0.3 \\ & \end{bmatrix} \end{array} A$$

$$\begin{array}{c} \textbf{TA1} \\ \textbf{TA2} \end{array} \begin{bmatrix} 0.27 & \\ -0.12 & 0.22 \end{bmatrix} B$$

Expression value (E)

$$\begin{array}{c} \\ \textbf{TF1} \\ \textbf{TF2} \end{array} \begin{array}{c} \mathbf{E_0} \\ \begin{bmatrix} 10 \\ 8 \end{bmatrix} \end{array} V_0$$

$$\begin{array}{c} \textbf{TA1} \\ \textbf{TA2} \end{array} \begin{bmatrix} 0.4 \\ 42 \end{bmatrix} W_0$$

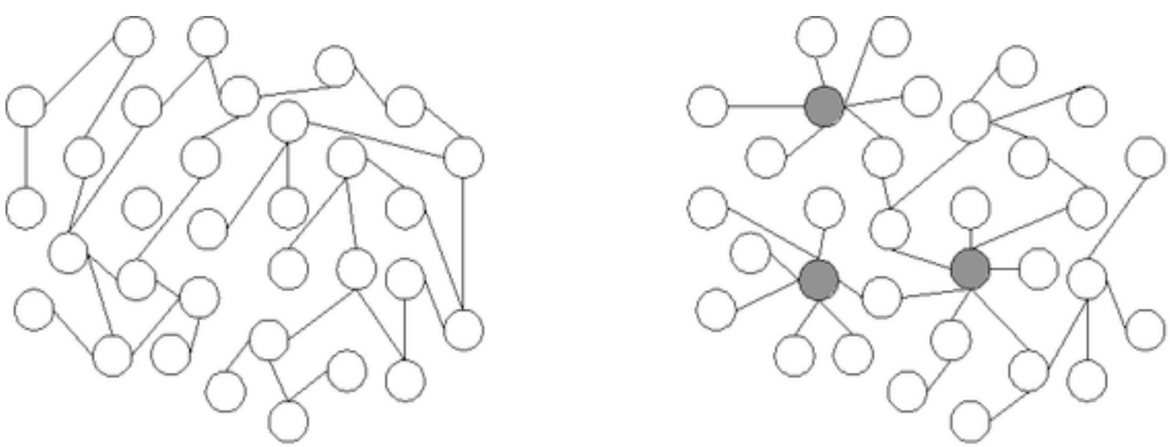
$$V(t+1) = V(t) + A.V(t)$$
$$W(t+1) = W(t) + B.V(t)$$

Carré et al, *npj Systems Biology & App* (2017)

“biological” constrains = Mathematical challenges.

-1) Chose Number of TF and TA (potentially big 10^4)
-2) Number of connections by Gene (TF or TA)
-3) Scale-Free properties (Connection degree follow a power law)
-4) Generated gene expression is following “normal distribution”
-5) The network as to generate stable gene expression.

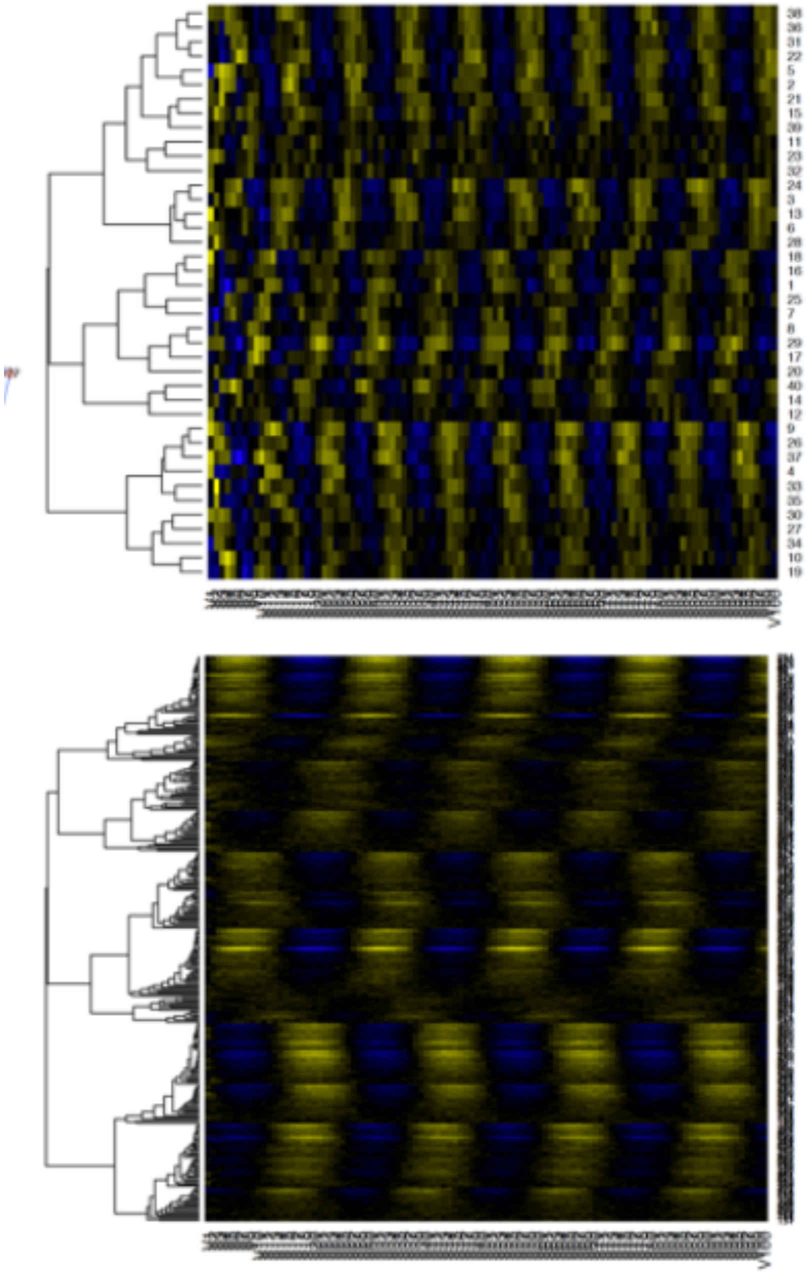
$$P(k) \sim k^{-\gamma}$$



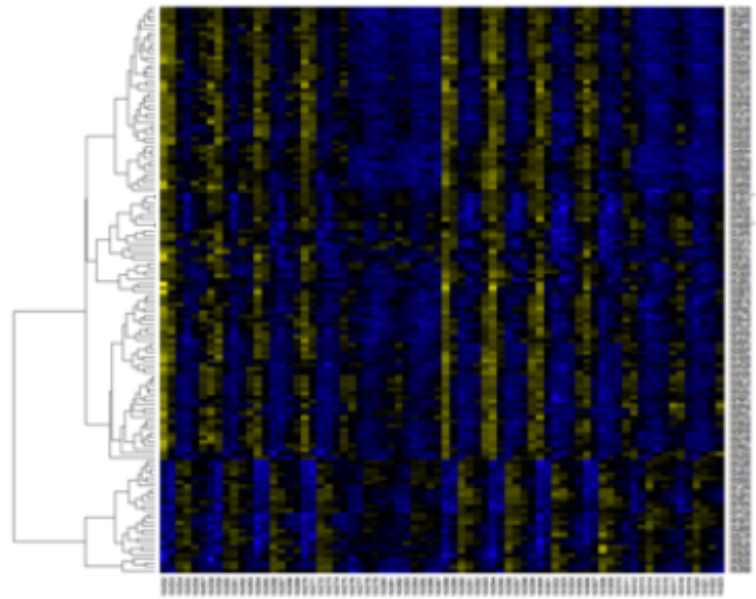
(a) Random network (b) Scale-free network

Carré et al, *npj Systems Biology & App* (2017)

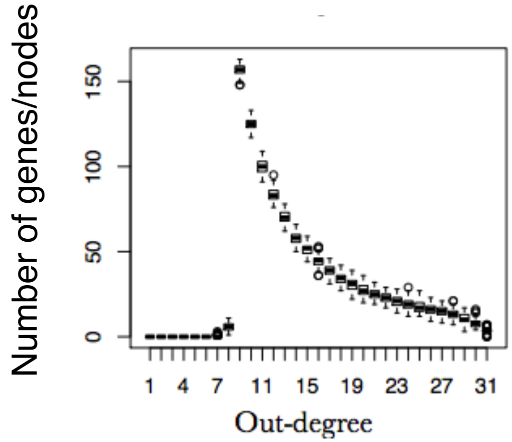
FRANK DATA

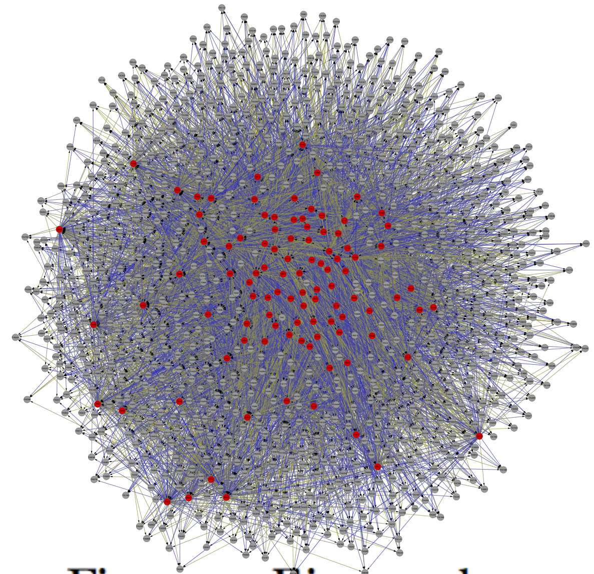


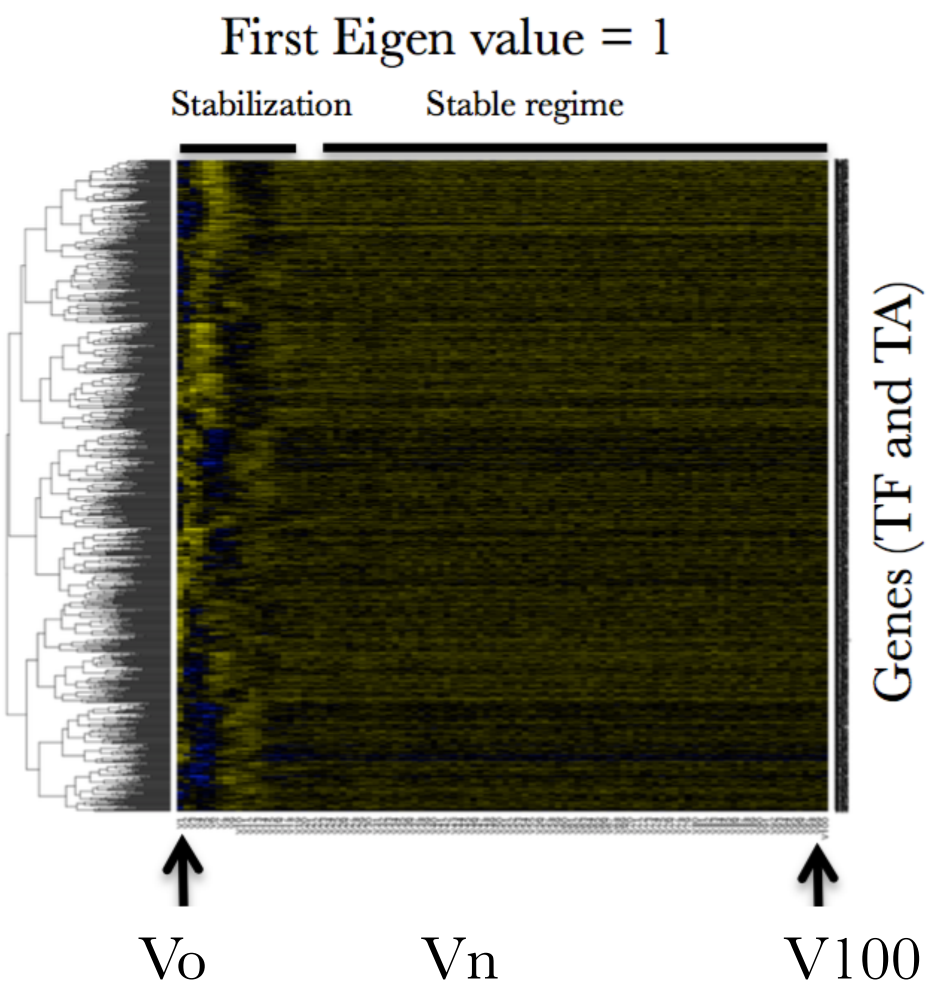
Affy DATA (Kay lab) Real data

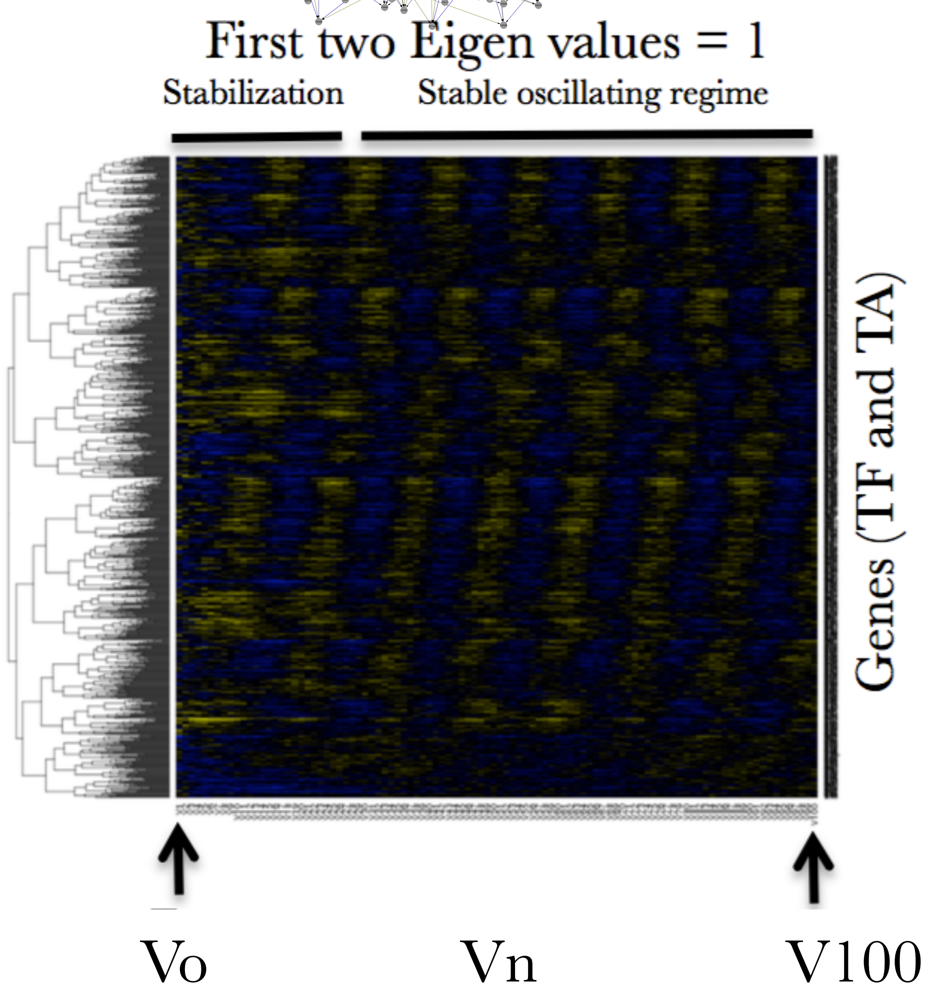


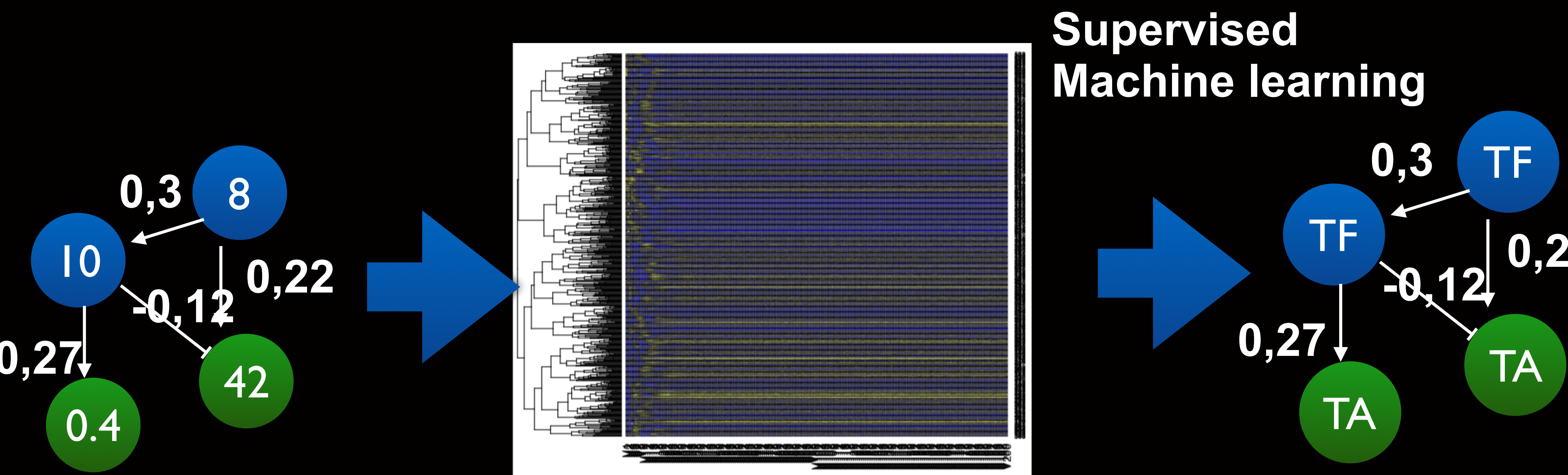
100 TF and 1000 TA

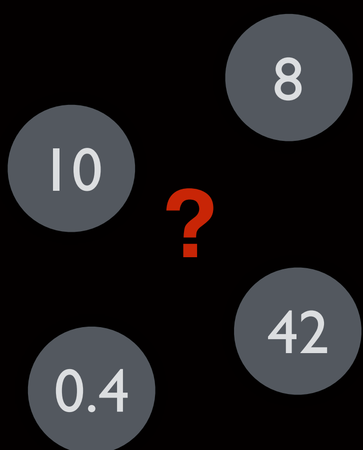


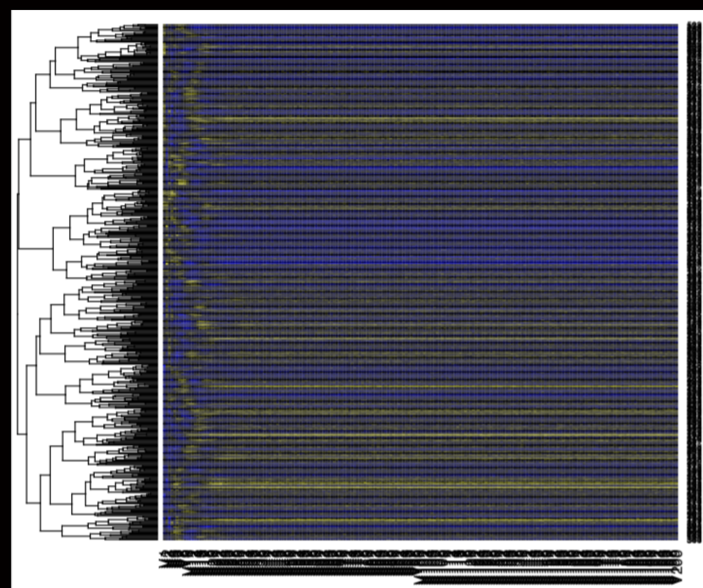




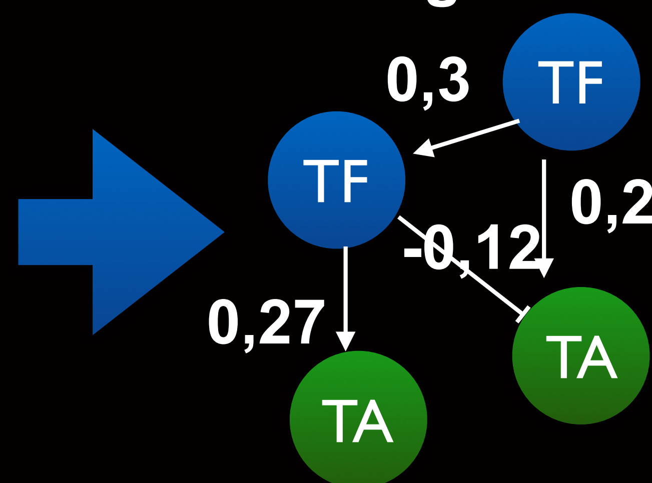


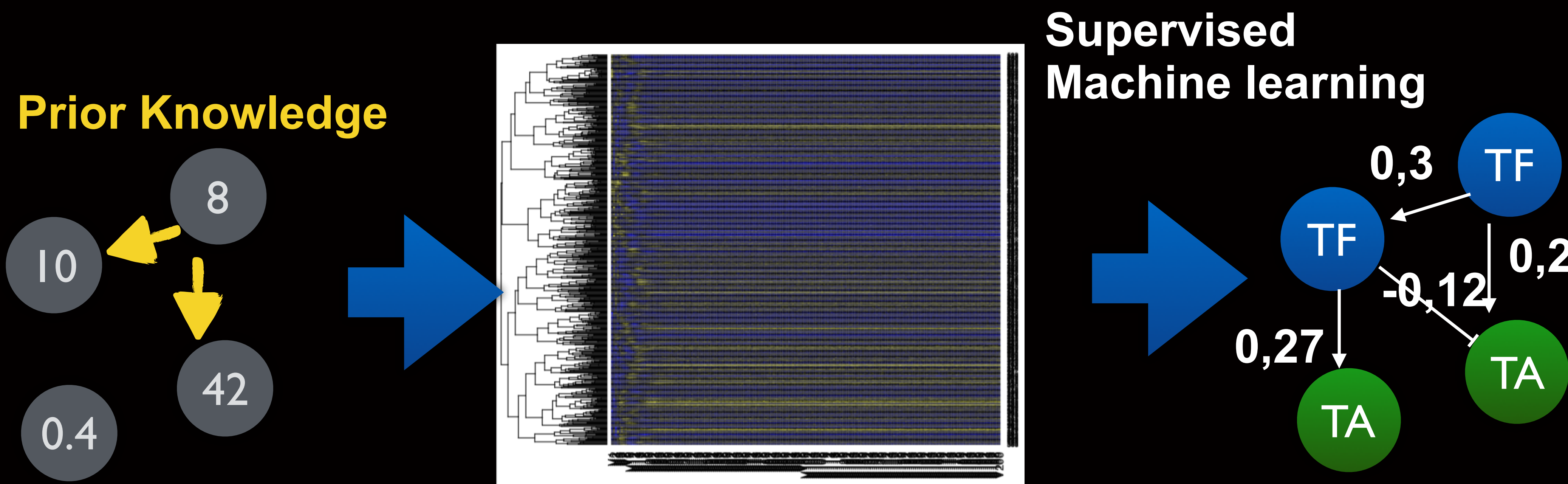

0,3
8
10
0,22
-0,12
0,27
0.4
42
Supervised
Machine learning
0,3
TF
TF
0,2
-0,12
0,27
TA
TA





Supervised Machine learning

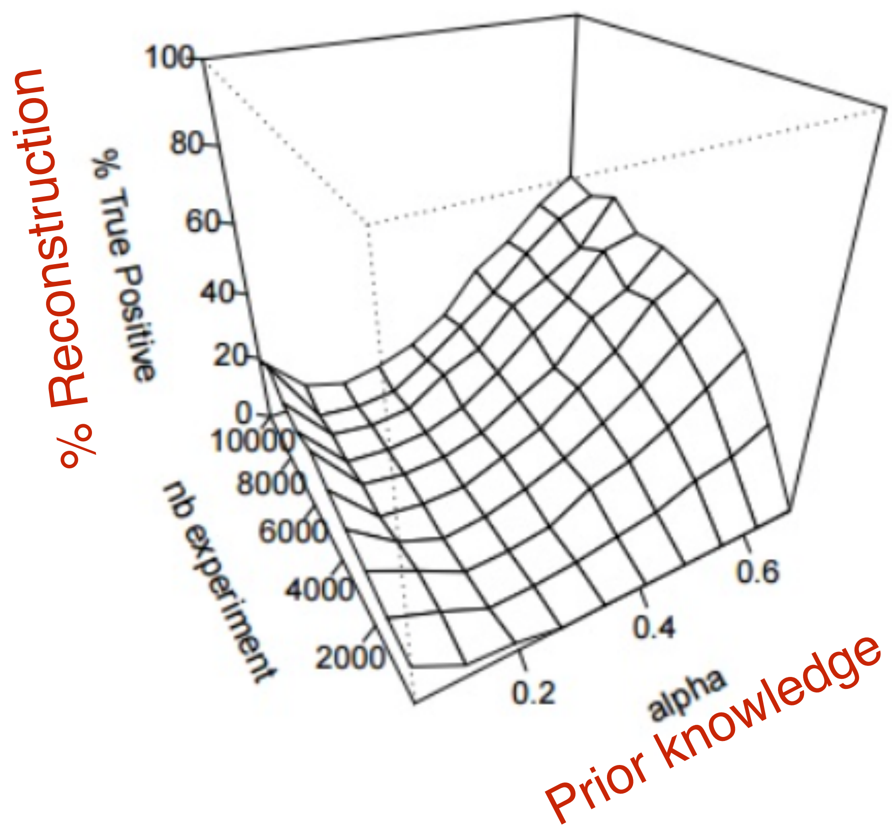




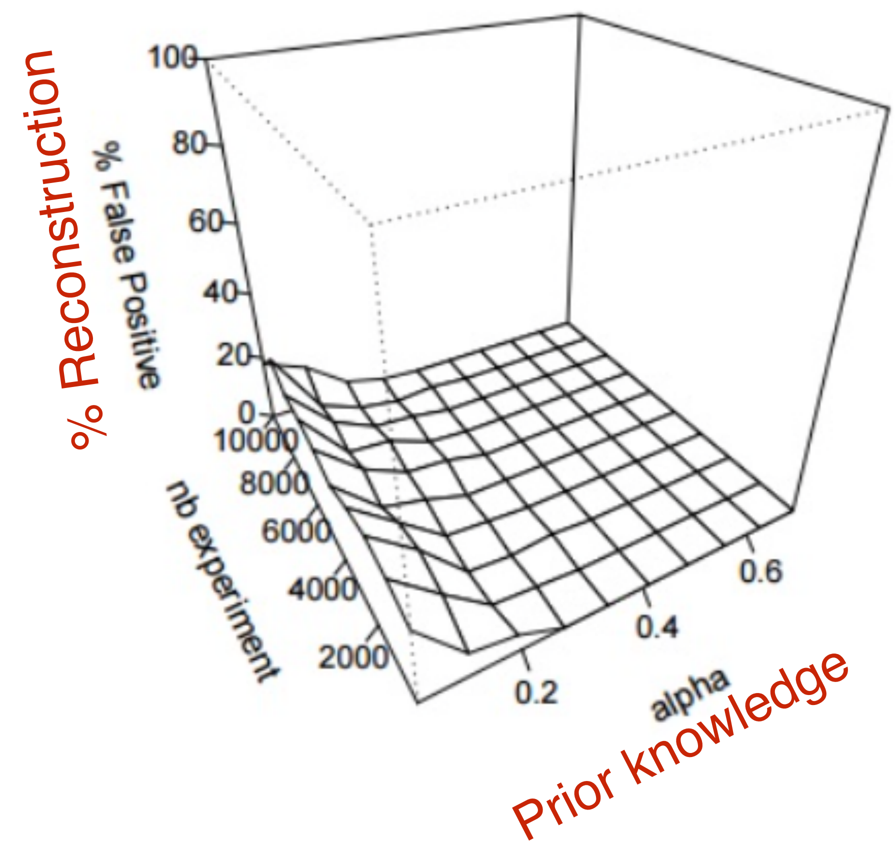
Prior Knowledge
8
10
42
0.4
Supervised
Machine learning
TF
TF
TA
TA
0,3
0,2
-0,12
0,27

Learning a 100 TF, 1000 TA GRN with SVM (Machine Learning Algorithm).

True Positives :
% unknown non-zero values
well predicted



False Positives :
% unknown zero values
predicted as non-zero



Carré et al, *npj Systems Biology & App* (2017)

Different Prior Structure

Network (N)

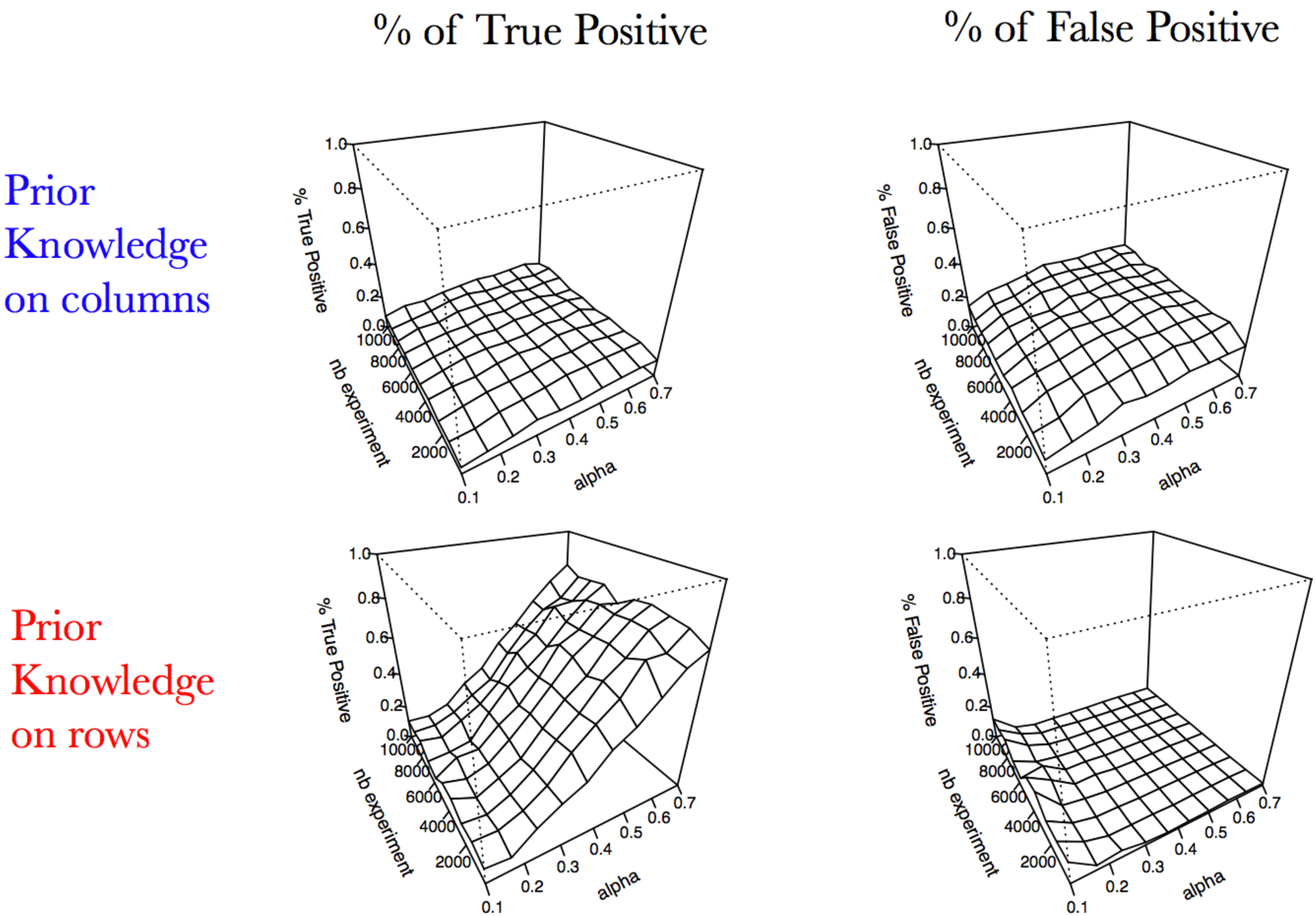
	TF1…	TFi…	TFn
TF1	?	-1	?
…			
TFi	1	0	-1
…			
TFn	?	0	?
TA1	?	1	?
…			
Tai	?	0	?
…			
TAn	0	-1	1

Prior Knowledge on TF action (columns)

Prior Knowledge on TA regulation (rows)

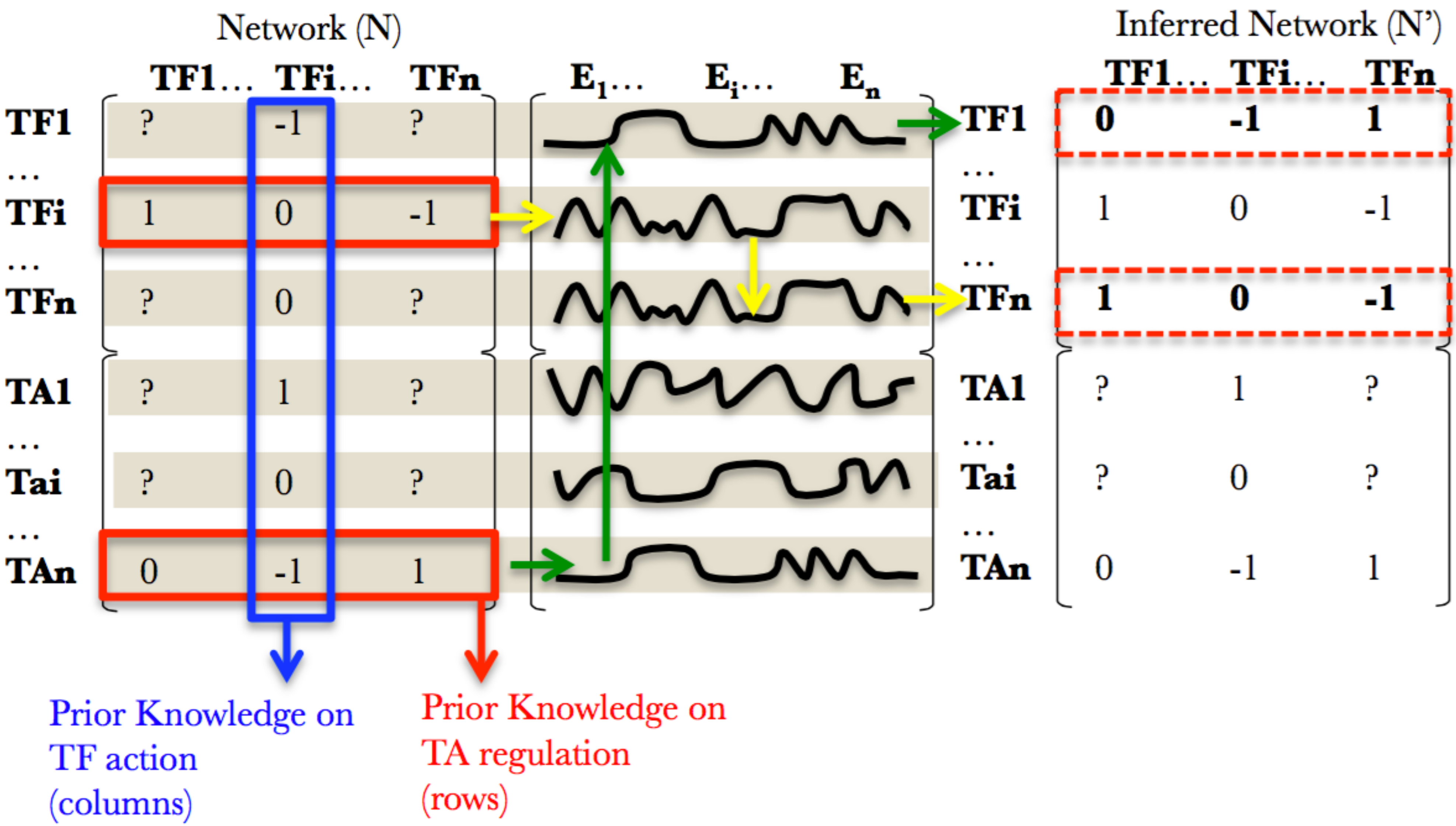
Carré et al, *npj Systems Biology & App* (2017)

TA oriented prior knowledge is predicted to be superior to supervise SVM machine learning procedures as compared to TF oriented techniques



Carré et al, *npj Systems Biology & App* (2017)

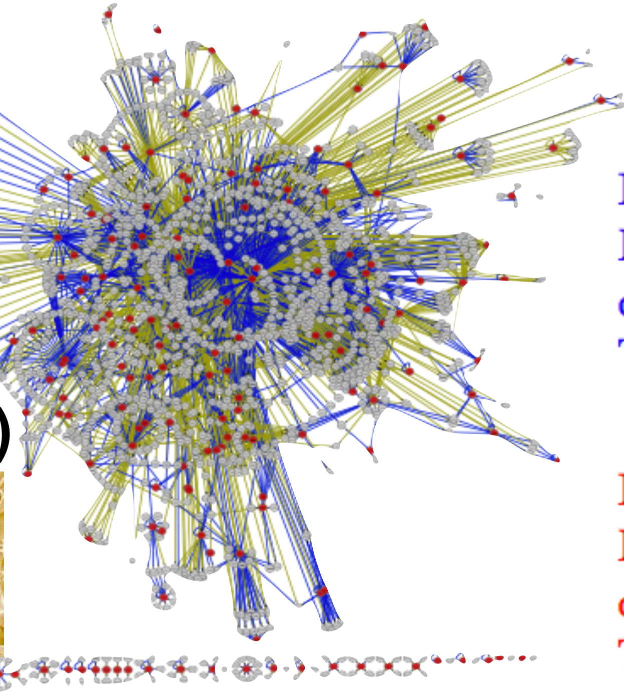
This is probably why ;)



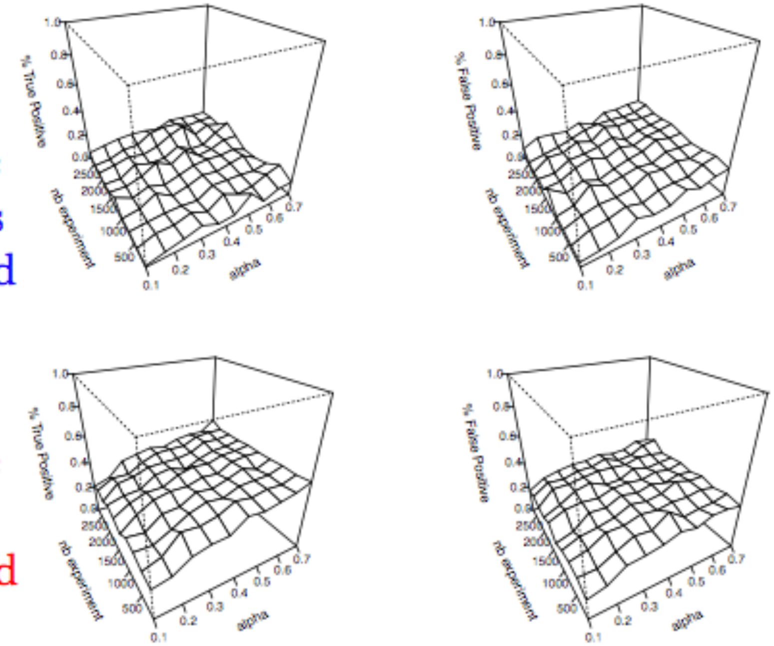
Carré et al, *npj Systems Biology & App* (2017)

Reviewers asked for proof on real data

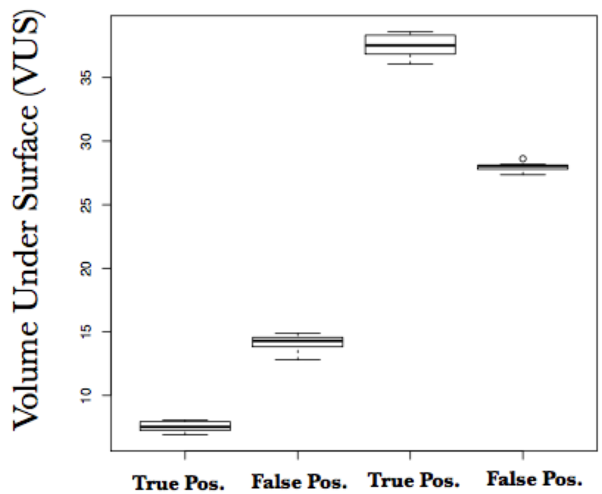
(here E.coli)



	% of True Positive	% of False Positive*
Prior Knowledge on columns TF-oriented		
Prior Knowledge on rows TA-oriented		



* This may contains true network connections not uncovered yet in the current version (9.3) of the E.coli Network.



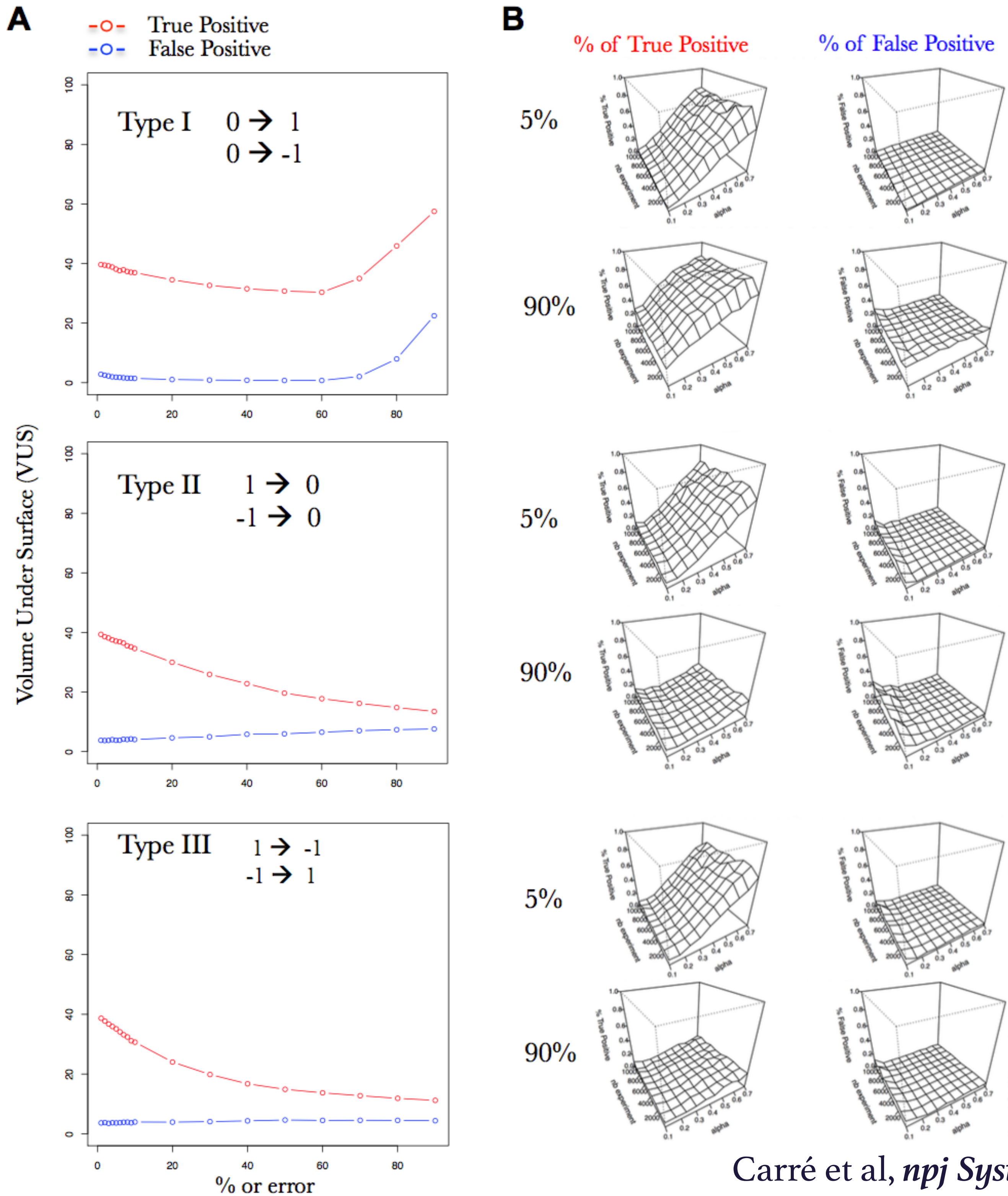
Prior Knowledge on columns TF-oriented

Prior Knowledge on rows TA-oriented

And it works!!!

Somehow demonstrating that we are in the good path.

Carré et al, *npj Systems Biology & App* (2017)



Carré et al, *npj Systems Biology & App* (2017)

Many more conclusions about biological system simulation and predictions.

-Gene expression dynamics…

-Learning resilient to noise…

All the math tricks fully detailed in the paper.

A useful algorithm available at : ***www.m2sb.org***

Carré et al, *npj Systems Biology & App* (2017)

Plan

1- Intro

2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

4- Is Regine cool?
(REGINE)

Plan

1- Intro

2- Computer modeling and experimental approaches converge towards the same conclusions concerning GRNs (FRANK)

3- Can we infer causality in static data? Maybe yes. (TRANSDETECT)

4- Is Regine cool? (REGINE)